

Docket No. RSW920010141US1

EXPRESS MAIL LABEL NO.: ET402935188US

DATE OF DEPOSIT: September 21, 2001

I hereby certify that this paper and fee are being deposited with the United States Postal Service Express Mail Post Office to Addressee service under 37 CFR §1.10 on the date indicated above and is addressed to the Assistant Commissioner of Patents, Washington, D.C. 20231.

Dianne Lane

NAME OF PERSON MAILING PAPER AND FEE

Dianne Lane

SIGNATURE OF PERSON MAILING PAPER AND FEE

INVENTORS: Steven M. Berman, Thirumale Niranjana, Srikanth Ramamurthy,
Sumanthkumar Sukumar

**METHOD AND APPARATUS FOR MINIMIZING INCONSISTENCY
BETWEEN DATA SOURCES IN A WEB CONTENT
DISTRIBUTION SYSTEM**

CROSS REFERENCE TO RELATED APPLICATIONS

5 The present invention is related to an application
entitled Method and Apparatus for Caching Subscribed and
Non-Subscribed Content in a Network Data Processing
System, serial no. _____, attorney docket no.
RSW920010142US1, filed even date hereof, assigned to the
10 same assignee, and incorporated herein by reference.

0960451.092101
F01260" F5409660

BACKGROUND OF THE INVENTION**1. Technical Field:**

The present invention relates generally to an improved data processing system, in particular to a method and apparatus for processing data. Still more particularly, the present invention provides a method, apparatus, and computer implemented instructions for distributing web content and minimizing inconsistencies between data sources.

10 2. Description of Related Art:

The Internet, also referred to as an "internetwork", is a set of computer networks, possibly dissimilar, joined together by means of gateways that handle data transfer and the conversion of messages from a protocol of the sending network to a protocol used by the receiving network. When capitalized, the term "Internet" refers to the collection of networks and gateways that use the TCP/IP suite of protocols.

The Internet has become a cultural fixture as a source of both information and entertainment. Many businesses are creating Internet sites as an integral part of their marketing efforts, informing consumers of the products or services offered by the business or providing other information seeking to engender brand loyalty. Many

09960451.092101
T01260" T5409660

Docket No. RSW920010141US1

federal, state, and local government agencies are also employing Internet sites for informational purposes, particularly agencies which must interact with virtually all segments of society such as the Internal Revenue

5 Service and secretaries of state. Providing informational guides and/or searchable databases of online public records may reduce operating costs. Further, the Internet is becoming increasingly popular as a medium for commercial transactions.

10 Currently, the most commonly employed method of transferring data over the Internet is to employ the World Wide Web environment, also called simply "the Web". Other Internet resources exist for transferring information, such as File Transfer Protocol (FTP) and Gopher, but have not
15 achieved the popularity of the Web. In the Web environment, servers and clients effect data transaction using the Hypertext Transfer Protocol (HTTP), a known protocol for handling the transfer of various data files (e.g., text, still graphic images, audio, motion video,
20 etc.). The information in various data files is formatted for presentation to a user by a standard page description language, the Hypertext Markup Language (HTML). In addition to basic presentation formatting, HTML allows developers to specify "links" to other Web resources
25 identified by a Uniform Resource Locator (URL). A URL is a special syntax identifier defining a communications path to

09960451-092101
TOT260-T5409660

Docket No. RSW920010141US1

specific information. Each logical block of information accessible to a client, called a "page" or a "Web page", is identified by a URL. The URL provides a universal, consistent method for finding and accessing this

5 information, not necessarily for the user, but mostly for the user's Web "browser". A browser is a program capable of submitting a request for information identified by an identifier, such as, for example, a URL. A user may enter a domain name through a graphical user interface (GUI) for
10 the browser to access a source of content. The domain name is automatically converted to the Internet Protocol (IP) address by a domain name system (DNS), which is a service that translates the symbolic name entered by the user into an IP address by looking up the domain name in a database.

15 The Internet also is widely used to transfer applications to users using browsers. With respect to commerce on the Web, individual consumers and business use the Web to purchase various goods and services. In offering goods and services, some companies offer goods and services
20 solely on the Web while others use the Web to extend their reach.

Content distribution systems are employed by businesses and entities delivering content, such as Web pages or files to users on the Internet. Currently,
25 content providers will set up elaborate server systems or other types of data sources to provide content to various

09960451-092101
T01250-T5409550

Docket No. RSW920010141US1

users. Web content distribution systems are those systems that are employed to distribute content to these servers and caches. This type of setup includes various nodes that act as sources of data. In this type of content distribution scheme, data from a primary or publishing node is propagated to all of the other nodes in the system. These types of systems require maintenance in addition to being expensive to put in place.

10 When a node within the system receives a notification that content is being propagated, the node pulls the data from a server or other data source and makes the data available to external clients requesting the data. In an ideal situation, accesses by clients are coordinated with the modification of the data at the various nodes in the system or a client always pulls data from a single node. In this situation, the data read by a single external client is guaranteed to be internally consistent.

20 Unfortunately, the ideal situation is currently unachievable because central coordination between external clients, nodes such as Web servers and caches, are not practical when scalability and performance are important. Further, different nodes may have dissimilar rates of data retrieval from Web servers and external clients cannot be blocked to ensure the node with the

09960451-092101
TOTAL 260

Docket No. RSW920010141US1

slowest connection to its data server becomes consistent with other nodes without a degradation of performance. Additionally, with the use of one or more load balancers between a client and a data source, a client may receive
5 the same data from two different servers depending on network conditions.

Therefore, it would be advantageous to have an improved method, apparatus, and computer implemented instructions for distributing content and minimizing
10 inconsistency between data sources.

SUMMARY OF THE INVENTION

The present invention provides a method, apparatus and computer implemented instructions for minimizing inconsistency between a set of data sources in a data
15 processing system. A first signal is sent indicating that new content is present for the set of data sources. The new content is transmitted to the set of data sources in which the new content is unavailable for distribution by the set of data sources until a second signal is
20 received by the set of data sources. The second signal is sent to the set of data sources if an acknowledgment is received from all of the sets of data sources.

09960451-092101

BRIEF DESCRIPTION OF THE DRAWINGS

The novel features believed characteristic of the invention are set forth in the appended claims. The invention itself, however, as well as a preferred mode of use, further objectives and advantages thereof, will best be understood by reference to the following detailed description of an illustrative embodiment when read in conjunction with the accompanying drawings, wherein:

Figure 1 is a network data processing system in accordance with a preferred embodiment of the present invention;

Figure 2 is a block diagram of a data processing system that may be implemented as a server in accordance with a preferred embodiment of the present invention;

Figure 3 is a diagram illustrating data flow in updating content at data sources in accordance with a preferred embodiment of the present invention;

Figure 4 is a flowchart of a process used for updating content in a content distribution system in accordance with a preferred embodiment of the present invention;

Figure 5 is a flowchart of a process used for updating content in a data source in accordance with a preferred embodiment of the present invention;

09960451-092101
TOT260-15409660

Docket No. RSW920010141US1

Figure 6 is a flowchart of a process used for initiating a contract for providing content distribution services in accordance with a preferred embodiment of the present invention; and

5 Figure 7 is a flowchart of a process used for billing a customer in accordance with a preferred embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

With reference now to the figures and in particular
10 to Figure 1, a network data processing system is depicted in accordance with a preferred embodiment of the present invention. Network data processing system 100 in this example includes network 102, which interconnects servers 104, 106, 108 and 110. These servers provide content to
15 clients, such as clients 112, 114, and 116, through network 102. In this example, network 102 takes the form of the Internet.

Servers 104-110 are servers within a Web content distribution system. This system also includes content
20 management and creator 118, which is connected to server 110 by local area network (LAN) 120. This Web content distribution system is also referred to as a content distribution framework and is an example of a system in

09960451-092101
FOI250-T5409660

Docket No. RSW920010141US1

which inconsistency between data and data sources is minimized, such as servers 104-108. In this example, server 110 functions as a primary publishing node while servers 104-108 serve as data sources to provide content to users making requests. Server 110 includes a master content distribution server and a master content distribution (CD) server process 122. Master content distribution server process 122 accepts notifications of new, deleted, or modified content from content management and creator 118. These notifications are propagated to servers 104-108, which then can invalidate or pull updated content from various sources. The content may be pulled from server 110 or from other sources. Typically, when a content publisher issues a notification to master CD server 122 in server 110, an identification of a staging server containing the content is made. Each of the servers pulling content includes a content distribution process (not shown), which will update content on a server when a notification is received.

This framework may be used to distribute multiple content types. For example, the framework may be used to move static content. Additionally, the framework may be used to publish or present documents on Web sites. In this instance, the framework will send notifications to the various nodes from the publishing node. The framework takes up the responsibility of updating the

00960451-092101
TOT260" T540960

Docket No. RSW920010141US1

various repositories. Next, the framework may be used to move applications to the nodes for distribution and use. Third, the framework may be used to manage cached dynamic content. Finally, the framework may be used to

5 distribute media files. Media files are similar to static pages. However, their large size requires a slightly different treatment. The transport mechanism in the framework may include mechanisms to pace the data distribution depending on factors such as the media type,

10 the bandwidth requirements, and available bandwidth.

The present invention provides a method, apparatus, and computer implemented instructions for managing content within this type of framework. In particular, the present invention provides a mechanism for minimizing

15 the window of inconsistency between data sources as well as describing a framework for providing content distribution to clients who create content, but do not necessarily desire to set up or maintain a content distribution system.

20 Content distribution services may be provided using this architecture by basing business contracts on guaranteeing a level of service. This level of service may include one or more of the following: bandwidth, storage, freshness or management. In these examples,

25 bandwidth is the certified distribution bandwidth between internal nodes and out to remote clients. Storage is the

00960451-092101
101260-15109600

Docket No. RSW920010141US1

amount of continuously available storage on current media. Freshness is the assurance that all content served will be up-to-date with respect to its origin.

Management is the provision of management tools to
5 manipulate the distribution parameters and locations.

With the terms of a contract in place to establish these parameters for service, the service provider and their customers both have a certain limitation in their ability to enforce the contract.

10 In these examples, content owners, the customers, may be required to establish at least one (edge) server with the content they wish to distribute available to the service provider with the content distribution system. At least one server is designated to handle the content
15 bundles that the owner wishes to distribute. Once the subscription mapping is in place, the content updates are automatically sent to the distribution network.

Monitoring nodes may be used to detect the freshness of the content being served and to report the success rate
20 of content updates to the administrators.

The framework may employ a content distribution system to migrate or replicate Internet content to remote servers according to a predetermined schedule, or other automatically generated criteria. A rule based system or
25 a dynamic use-analysis feedback system can automatically replicate the currently "hot" assets to more and more

09960451-092101
T01260-T5409560

Docket No. RSW920010141US1

external caches if they reside within the system's administrative domain or within the domain of a cooperating CDSP. Whenever content becomes "important" for either popularity or other reasons, it can be
5 migrated out to the replica sites automatically under program control.

Referring to **Figure 2**, a block diagram of a data processing system that may be implemented as a server, such as server 104 in **Figure 1**, is depicted in accordance
10 with a preferred embodiment of the present invention. Data processing system 200 may be a symmetric multiprocessor (SMP) system including a plurality of processors 202 and 204 connected to system bus 206. Alternatively, a single processor system may be employed.
15 Also connected to system bus 206 is memory controller/cache 208, which provides an interface to local memory 209. I/O bus bridge 210 is connected to system bus 206 and provides an interface to I/O bus 212. Memory controller/cache 208 and I/O bus bridge 210 may be
20 integrated as depicted.

Peripheral component interconnect (PCI) bus bridge 214 connected to I/O bus 212 provides an interface to PCI local bus 216. A number of modems may be connected to PCI local bus 216. Typical PCI bus implementations will
25 support four PCI expansion slots or add-in connectors. Communications links to clients 108-112 in **Figure 1** may be

09960451.092101
TOT260" T5H09660

Docket No. RSW920010141US1

provided through modem 218 and network adapter 220
connected to PCI local bus 216 through add-in boards.

Additional PCI bus bridges 222 and 224 provide
interfaces for additional PCI local buses 226 and 228,
5 from which additional modems or network adapters may be
supported. In this manner, data processing system 200
allows connections to multiple network computers. A
memory-mapped graphics adapter 230 and hard disk 232 may
also be connected to I/O bus 212 as depicted, either
10 directly or indirectly.

Those of ordinary skill in the art will appreciate
that the hardware depicted in **Figure 2** may vary. For
example, other peripheral devices, such as optical disk
drives and the like, also may be used in addition to or in
15 place of the hardware depicted. The depicted example is
not meant to imply architectural limitations with respect
to the present invention.

The data processing system depicted in **Figure 2** may
be, for example, an IBM e-Server pSeries system, a
20 product of International Business Machines Corporation in
Armonk, New York, running the Advanced Interactive
Executive (AIX) operating system or LINUX operating
system.

Within the updating and distribution of content the
25 present invention also includes a mechanism for
minimizing windows of inconsistency between the different

09960451-092101
F01260-15409660

Docket No. RSW920010141US1

data sources. In particular, the present invention provides a method, apparatus, and computer implemented instructions for minimizing the window of inconsistency between data sources, such as, web caches or web servers, by distributing notifications of updates in a two-phase manner. The two phases ensure that fresh content is made live at roughly the same time at all caches and servers. In these examples, the mechanism is implemented in a content distribution system that provides a notification, which results in a pulling of the data to the data sources. Of course, this mechanism may also be implemented in systems that push data to data sources. In each case, the data is made available to requestors when all of the data sources contain the updated content.

15 The mechanism of the present invention is performed without requiring a central coordinator to arbitrate client browser requests and notifications. In other words, client requests go straight to the caches and servers, and the clients see consistent data across the nodes. This process is performed without requiring that external clients wait until the nodes become consistent with each other. The mechanism of the present invention will disconnect nodes that are unable to update content without requiring other nodes to roll back or use old content. During the whole two-phase process, a node can serve out old content, maintaining high availability.

Docket No. RSW920010141US1

With reference now to **Figure 3**, a diagram illustrating data flow in updating content at data sources is depicted in accordance with a preferred embodiment of the present invention. In this example, content at Web server 300 and Web server 302 is updated from content located at originating Web server 304.

These servers are servers in a Web content distribution system such as that illustrated in **Figure 1**. Web server 300 includes temporary storage 306 and available content 308. Similarly, Web server 302 includes temporary storage 310 and available content 312.

When a user requests content from a client, such as client 314, the request is typically made from a browser, such as browser 316. The request may be routed to either Web server 300 or Web server 302 through a load balancing system. If Web server 300 receives the request, the content returned to client 314 is returned from content in available content 308. This content may be, for example, a Web page or an audio file. If the request is routed to Web server 302, the content is returned to client 314 from content in available content 312. In either case, the content is identical.

At some point, changes to the content in available content 308 and available content 312 may be made. For example, a new Web page may be added, a Web page may be modified, or a Web page may be deleted from the content.

Docket No. RSW920010141US1

The initiation of this process occurs when a signal indicating that content is to be updated is received by Web server 300 and Web server 302. This signal is received from originating Web server 304 in this example.

5 In these examples, Web server 300 and Web server 302 pull the content from originating Web server 304. The content is stored in temporary storage 306 and temporary storage 310 during the pull process. When Web server 300 receives all of the new content, this Web server sends an
10 acknowledgment signal back to originating Web server 304. Similarly, Web server 302 will transmit an acknowledgment signal to originating Web server 304 when Web server 302 has pulled all of the new content. The completion of the pulling of new content may occur at different times in
15 Web server 300 and Web server 302 depending on the various network conditions, such as available bandwidth, network traffic, and the number of hops to originating Web server 304.

This content is not made available to clients until
20 a second signal is received from originating Web server 304 indicating that the content is to be published or made available in response to request from clients. During this time, the content in available content 308 and available content 312 is used to reply to requests
25 from clients.

10141US1-092260

Docket No. RSW920010141US1

In this manner, the content available at Web server 300 and Web server 302 is consistent. When the second signal is received, the content from temporary storage is placed into available content at each Web server. In this manner, the window of inconsistency between different servers is minimal. With the second signal being sent to Web server 300 and Web server 302 at the same time, the window of inconsistency between these two nodes is reduced significantly even if these two nodes have very different connection speeds. As a result, the content is made available at around the same time.

If Web server 300 does not pull all of the content from originating Web server 304 or is unable to return an acknowledgment signal, originating Web server 304 will disconnect Web server 300 and will send the second signal to Web server 302 after some period of time. This period of time is selected as one indicating that a server is malfunctioning or may be based on other factors, such as performance. This minimizing the window of inconsistency between data sources may be offered as part of a service for which a client is charged or billed.

Turning next to Figure 4, a flowchart of a process used for updating content in a content distribution system is depicted in accordance with a preferred embodiment of the present invention. The process illustrated in Figure 4 may be implemented in an

0960451-092101

Docket No. RSW920010141US1

originating Web server, such as originating Web server 304 in Figure 3.

5 The process begins by sending content notification to nodes in a group (step 400). A determination is made as to whether an acknowledgment has been received (step 402). An acknowledgment is returned from a node to the originating Web server when all of the content has been propagated to the node. If an acknowledgment has been received, a determination is then made as to whether an acknowledgment has been received from all nodes in the group (step 404). This step is used to determine if all of the nodes have received the new content. If an acknowledgment has been received by all nodes in the group, published messages are sent to all nodes in the group (step 406) with the process terminating thereafter. 10 The published message causes the nodes to make the new content available in response to requests from users. 15

With reference again to step 404, if an acknowledgment has not been received by nodes in the group, a determination is made as to whether a timeout has occurred (step 408). The timeout period is set as a period of time after which an assumption is made that a node is malfunctioning or network conditions have made it impossible to return an acknowledgment. If a timeout has not occurred, the process returns to step 402. 20 25

Otherwise, the nodes from which an acknowledgment has not

00960451-092101
101220-T540960

Docket No. RSW920010141US1

been received are removed from the group (step 410) and the process proceeds to step 406 as described above. With reference again to step 402, if an acknowledgment is not received, the process also proceeds to step 408, as
5 described above.

Turning next to **Figure 5**, a flowchart of a process used for updating content in a data source is depicted in accordance with a preferred embodiment of the present invention. The process illustrated in **Figure 5** may be
10 implemented in a data source, such as Web server 300 in **Figure 3**.

The process begins by receiving a new content message (step 500). Content is received (step 502). The content may be received by the data source pulling the
15 content or from a push from a server originating the new content. A determination is then made as to whether all content has been received (step 504). If all content has been received, an acknowledgment is sent back to the server initiating the update (step 506). The process
20 then waits for a publish message (step 508). After receiving the publish message, the new content is made available to requests (step 510) with the process terminating thereafter. During the time when the new content is unavailable in response to requests, the old
25 content is used to respond to these requests.

Docket No. RSW920010141US1

With reference now to **Figure 6**, a flowchart of a process used for initiating a contract for providing content distribution services is depicted in accordance with a preferred embodiment of the present invention.

- 5 The process illustrated in **Figure 6** may be implemented in a network data processing system, such as network data processing system 100 in **Figure 1**.

10 The process begins by receiving a request to host content from a customer (step 600). This request may be made through selection of a link in a Web page. Contract terms are sent to the customer (step 602). These terms may include, for example, the quality of service that may be guaranteed, an identification of resources made available to the client, billing rates, content to be
15 provided by the client, prohibited content, disclaimers, and other terms.

A determination is made as to whether the customer accepts the terms of the contract (step 604). If the customer rejects the terms of the contract, the process
20 terminates. On the other hand, if the customer accepts the terms of the contract, customer information is requested (step 606). This customer information may include a user ID, a password, an IP address of a server from which the client will originate content, a billing
25 address, and other contact information. A response is received from the customer (step 608), and the customer

0960451-092101
T01260-1540960

Docket No. RSW920010141US1

is set up to host content (step 610) with the process terminating thereafter.

Turning next to **Figure 7**, a flowchart of a process used for billing a customer is depicted in accordance with a preferred embodiment of the present invention. The process illustrated in **Figure 7** may be implemented in a network data processing system, such as network data processing system 100 in **Figure 1**.

The process begins by identifying an unprocessed customer from a database (step 700). Next, a billing structure for the customer is retrieved (step 702). Then, a bill is generated for content service using the billing structure (step 704) with the process terminating thereafter.

Thus, the present invention provides a method, apparatus, and computer implemented instructions for minimizing a window of inconsistency between data sources in a web content distribution system. The mechanism of the present invention minimizes the window by distributing content to data sources, but not allowing the data sources to publish or make the content available until all of the data sources have received the content. The mechanism of the present invention also allows for generating revenues for a content distribution system owner by allowing the content distribution system owner to provide services including the distribution of content

09960451.092101
TOT 360" T5409660

Docket No. RSW920010141US1

to data sources and minimizing windows of inconsistency
between data sources by billing clients for these
services. In this manner, customers may have content
published without having to incur the expenses of setting
5 up or maintaining a content distribution system.

It is important to note that while the present
invention has been described in the context of a fully
functioning data processing system, those of ordinary
skill in the art will appreciate that the processes of
10 the present invention are capable of being distributed in
the form of a computer readable medium of instructions
and a variety of forms and that the present invention
applies equally regardless of the particular type of
signal bearing media actually used to carry out the
15 distribution. Examples of computer readable media
include recordable-type media, such as a floppy disk, a
hard disk drive, a RAM, CD-ROMs, DVD-ROMs, and
transmission-type media, such as digital and analog
communications links, wired or wireless communications
20 links using transmission forms, such as, for example,
radio frequency and light wave transmissions. The
computer readable media may take the form of coded
formats that are decoded for actual use in a particular
data processing system.

25 The description of the present invention has been
presented for purposes of illustration and description,

09960451-092101
TOTAL 260 TST09660

Docket No. RSW920010141US1

and is not intended to be exhaustive or limited to the invention in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art. The embodiment was chosen and described in
5 order to best explain the principles of the invention, the practical application, and to enable others of ordinary skill in the art to understand the invention for various embodiments with various modifications as are suited to the particular use contemplated.

09960451-092101
TOT260"TS409660